Alternative Approach for Environmental Education Evaluations: Pilot Attempt to Utilize Camera and Sensor Data

Ryo SAKURAI,* Kazuhiko W. NAKAMURA,† Kazunari HARUTA,‡ Kosaku HASHIMOTO,§ Yuya NAKATA,¶ Takayuki NAKATA**

Abstract: One of the limitations of traditional environmental education (EE) evaluation studies is its dependence on subjective judgment in terms of the program outcomes. We worked with information systems engineers to create a camera and sensor device for the participants to wear during an EE tour. The study determined whether we could objectively evaluate participants' reactions during a guided tour and procure data to understand the effectiveness of the program. The results showed a significant correlation between scores obtained through camera sensors and observation after excluding outliers, implying that such devices could be a potential future substitute for human observation. Camera/sensor data provide detailed and objective information related to participants' physical reaction during the program (e.g., the moment participants started to lose attention), to the environmental educators as well as tour guides, and by utilizing such data, practitioners could design and conduct more effective environmental education and communication programs.

Keywords: *attachable camera device, environmental education, human observation, junior high school students, interview, objective evaluation*

1. Introduction

The importance of environmental education (EE) for younger generations and general citizens has been recognized as indispensable for establishing a sustainable society where people are knowledgeable and willing to take action for conservation (Kollmuss and Agyeman 2002; Jacobson et al. 2007). One of the limitations of traditional EE evaluation studies is that they mostly depend on subjective judgment in terms of the program outcomes (Camargo and Shavelson 2009). Reviews of EE research on evaluations revealed that most studies used a subjective evaluation approach where

* Associate Professor, College of Policy Science, Ritsumeikan University
† Full-time Lecturer, Graduate School of Frontier Sciences, The University of Tokyo
‡ Graduate Student, Graduate School of Engineering, Information Systems Engineering, Toyama Prefectural University
§ Graduate Student, Graduate School of Engineering, Information Systems Engineering, Toyama Prefectural University
¶ Graduate Student, Graduate School of Engineering, Information Systems Engineering, Toyama Prefectural University
** Associate Professor, Information Systems Engineering, Toyama Prefectural University
Email: * ryosak@fc.ritsumei.ac.jp
† k_nakamura@edu.k.u-tokyo.ac.jp
** nakata@pu-toyama.ac.jp

Received on September 27, 2022, accepted after peer reviews on February 7, 2023; published online: June 6, 2023.

Asia-Japan Research Academic Bulletin, 2023
ONLINE ISSN 2435-306X, Vol.4, 32

researchers conducted surveys or interviews and respondents provided subjective answers (Zint 2013; Kormos and Gifford 2014).

On the one hand, surveys and interviews could provide fruitful information and insights on the effectiveness of the program, if research is conducted in a rigorous manner (Vaske 2008); however, these methods have shortcomings. During EE programs such as guided tours in nature or in-class lectures on environmental issues, participants' concentration levels vary. Subjective evaluation can cause various biases such as social desirability bias where respondents provide answers that they think are socially desirable (Lusk and Norwood 2009). For example, participants could answer that they learned a lot from the program even if that was not the case. There are factors that objective evaluation could identify. For example, whether certain people engaged in certain behaviors (e.g., recycling) could be observed easily, and therefore, researchers could objectively judge if such behaviors occurred or not. On the other hand, factors such as a change in awareness or motivation cannot be observed from outside which make objective evaluation challenging. However, we believe that even such cognitive factors could be and should be objectively evaluated through utilizing technologies. In this study, assuming that certain cognitive factors (e.g., motivation, level of focus) are reflected in a participant's physical reaction, we tried to identify such a reaction through using the data obtained from a wearable camera.

The aim of our study is to objectively evaluate and monitor participants' reactions during a guided tour and procure data that can help us understand the effectiveness of such programs. This paper presents the results of our pilot study. Our main research question was how much camera/sensor data could substitute conventional observation, survey or interview data, if it actually could. We aimed to answer this question by evaluating the data obtained through such an approach. [In this paper, "observation" refers to the observation of participants by researchers].

2. Previous Studies on Objective and Subjective Evaluation

Subjective and objective evaluation could vary depending on the field of study. Ali and Patnaik (2018) measured respondents' subjective perceptions of thermally comfortable open spaces and compared them with values of the physiologically equivalent temperature [objective index] of the place, including temperature and humidity. Both objective and subjective evaluations showed that parks with tree canopy density are more comfortable than market lanes without trees (Ali and Patnaik 2018).

Lee et al. (2018) evaluated the effect of situated teaching in nurse education and compared students' subjective empathy scores with objective structure clinical examination where teachers and patients evaluated the actual behavior of students. Results showed that subjective judgment did not significantly correlate with the objective evaluation scores observed by teachers and patients.

Pourmand (2012) evaluated wideband speech quality by comparing subjective evaluation (ratings of speech quality by listeners) and auditory-based objective metrics such as loudness pattern distortion, and revealed that such objective metrics were effective in predicting subjective impressions of audio sound quality.

Velarde et al. (2007) reviewed environmental psychology studies on the effects of visible landscape on health and found that while evaluation measures ranged from subjective (e.g., self-reports of emotional state) to objective (e.g., blood pressure, behavioral observation) approaches, both methods seemed to be valid in terms of identifying such effects.

While the majority of evaluation studies conducted in the field of EE used subjective evaluation (e.g., participants' survey) (Zint 2013), limited studies have been conducted to directly measure behavior instead of relying on self-reports. A classical study conducted by Asch and Shore

(1975) revealed on the basis of the observation that children exposed to an EE program were more likely engaged in conservation behavior than those in a control group. More recently, Camargo and Shavelson (2009) proposed observable actions that researchers could directly measure such as litter reduction behavior to effectively evaluate EE programs. The significance of objective evaluation is that it could obtain information on what really happened (in this case, actual observed behavior). By utilizing both subjective and objective evaluation data, one can understand if these data correspond to imply the potential impacts of the program.

On the contrary, no known research has been conducted in the field of EE that has objectively evaluated people's reaction and compared it with participants' subjective observation. While the limitation of only depending on subjective evaluation to assess the effects of EE program has been continuously pointed out by researchers and practitioners (Camargo and Shavelson 2009), developing evaluation tools that do not depend on human judgment is challenging as it goes beyond the expertise of most traditional EE researchers (e.g., social scientists or ecologists). Meanwhile, collaboration with experts in other fields (e.g., engineering) could reveal original and alternative evaluation types. Similar to the study by Ali and Patnaik (2018) and Pourmand (2012), we aimed to develop evaluation tools to systematically monitor the reaction of EE program participants and compared the results with the traditional EE evaluation approach. Contrary to previous studies that looked at environmental behavior (e.g., Asch and Shore 1975, Camargo and Shavelson 2009), we focused on the physical reactions of the participants during the program, and therefore, this research is categorized as program process evaluation rather than impact evaluation (Rossi et al. 2004).

3. Method

(1) Construction of Camera Device and Related Studies

Through collaboration with information systems engineers, we developed a camera device (Figure 1) that participants of EE programs could attach to their heads.

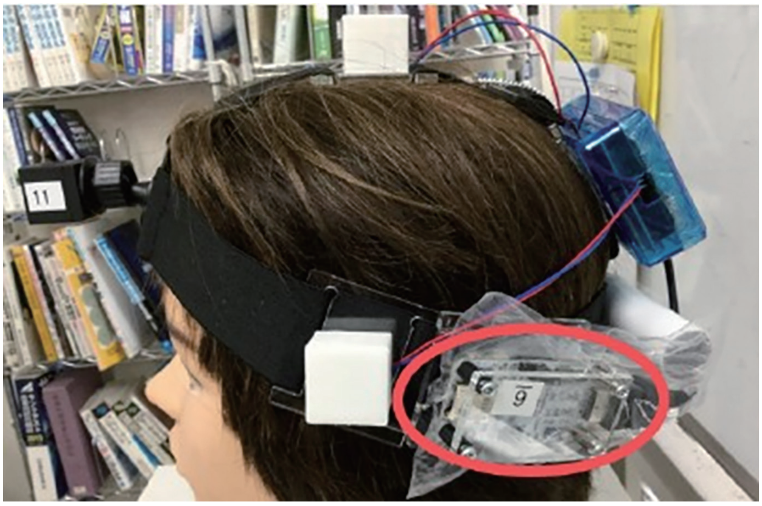

Figure 1. A camera device used for the research. Camera and accelerometer (covered by red circle) are attached to obtain velocity and angle of participants' head movements. (The authors)

There are studies that have measured the head or eye movement of participants by using wearable devices, especially in the field of engineering [e.g., the development of a new algorithm for measuring eye movements (Rothkopf and Pelz 2004) and testing individuals' remote interactions by attaching omnidirectional wearable cameras (Kasahara et al. 2017)]. Onishi et al. (2016) monitored participants' interest in the educational program through sensors attached to their heads as well as through expert observation and found that their levels of interest in the topic could be ascertained by their head movements.

Based on this research, we predicted that the participants' concentration on the program content or the guide's talk will be reflected in their head movements. Whether participants are looking at the tour guide while they talk or at objects pointed out by the guide can be tracked from participants' eye and head movements. Meanwhile, the validity of such objective evaluation needs to be verified by comparing it to a variety of conventional subjective evaluation methods (e.g., interviews). However, none of the studies mentioned here have used such methods of evaluation. Therefore, we utilized mixed methods covering a variety of subjective methods (observation, knowledge test, interview, and survey) to test our hypothesis that objective evaluation could be a substitute for conventional evaluation methods.

In this study, a portable device was created using products that are easily available in stores. A camera and sensor were attached to a headband. This device was pilot tested to confirm that the recorded video and sensory data could be used to obtain the velocity and angle of participants' head movements (Haruta et al. 2020). Following the previous study (Onishi et al. 2016), an average of a second was considered as one data point. We focused on that moment in the tour when the guide asked the participants to look at certain objects. Assuming that participants who quickly moved their heads in the direction of the object were more focused on the guide's talk, we calculated the individual scores of the participants by measuring the time before they moved their head in the appropriate direction. The study was conducted in four steps to measure the participants' concentration, which could ultimately demonstrate the effectiveness of the EE program.

(2) Procedure

First, we analyzed the video and sensor data to record how quickly students moved their heads in the direction the guide pointed.

Second, we observed participants' reactions and facial expressions during the tour. A researcher scored each student's reaction by observing the direction of their gaze (they scored 5 if they looked at the objects and 1 if they did not) as well as their focus on the guide's talk (they scored 5 if they looked very focused and 1 if they were not focused at all). Students' levels of concentration were measured every 30 seconds.

Third, the students were asked in an interview to state what they remembered about the program and score themselves (out of 100) in terms of how much attention they paid to the guide's talk and their satisfaction with the tour.

Finally, a survey was conducted to understand how the students scored themselves on their focus as well as their interest in the tour program. For this purpose, a 5-point Likert scale was used with a score of 5 if they were very focused, and 1 if they were not at all focused.

(3) Targeted Environmental Education Program

The EE guided tour was conducted at Shiga Kogen, a highland area located in the Joshin'etsukogen National Park of Japan. Guided nature tours are conducted here for students from all over the country, hence the place seemed appropriate for this research. The sample of this study comprised students from a junior high school in Ibaraki Prefecture (located in eastern Japan). All the second-year students (n = 76) joined the guided tour as part of a class activity. Students were divided into eight groups (each group having eight to ten students) and the professional guides (interpreters) took each group for a walk through the mountains. [In this study, we use the term environmental educators and guides as similar to the definition of interpreters; those who increase people's interest towards and understanding of the natural and cultural environment through communication (Nishimura 2016).] All 76 students wore the head mount cameras as well as the sensor devices and their facial expressions were observed and recorded. Eight students were randomly picked (one from each group) by choosing random numbers from students' list, and their facial expressions were observed during the tour. As there was only one researcher who could engage in observation, focusing on one student at a time, a total of eight students (one from each group) was observed. Interviews with the eight students were conducted after the tour, but all seventy-six students participated in the survey. From among the eight students selected for observation and interview, data from the camera/sensor of one student could not be obtained due to technical reasons, hence we deleted the data of that student. Finally, scores of seven students for four types of study were compared.

Before starting the study, we explained the research objectives to the teachers and students and informed them that all data obtained would be kept confidential and that students' participation would be voluntary. After obtaining informed consent from all the teachers and students, we started our research.

4. Results and Discussion

(1) Descriptive Results of Each Method

The students' head movements per second were recorded. Students whose speed of head movement toward the particular direction surpassed the threshold (see Appendix 1 for an explanation of students' head movement scores) were judged that their head movements were detected (Figure 2). Sensor data for that particular moment was extracted when a large number of students moved their heads at the same time (Figure 3).

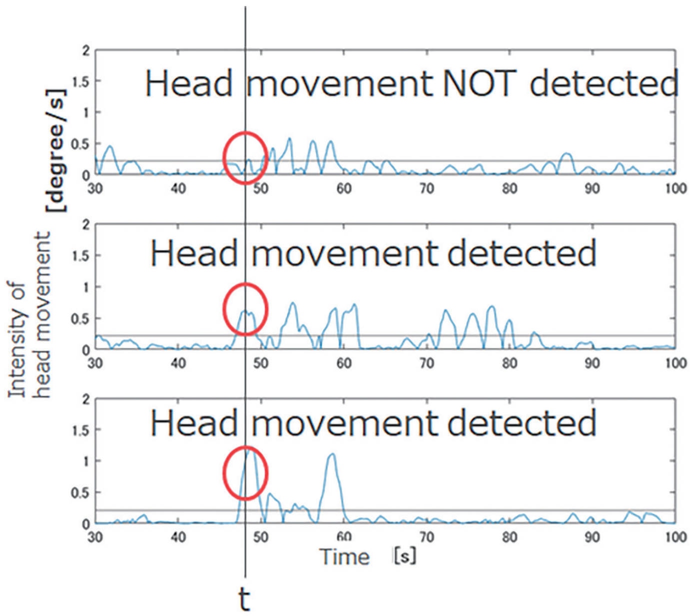

Figure 2. Movement of students' heads per second based on the camera-sensor data. Following the calculated standardized scores, two students' head movements were detected while another student was not at time t. [The figure was created by the authors based on Haruta et al.'s study (2020).]

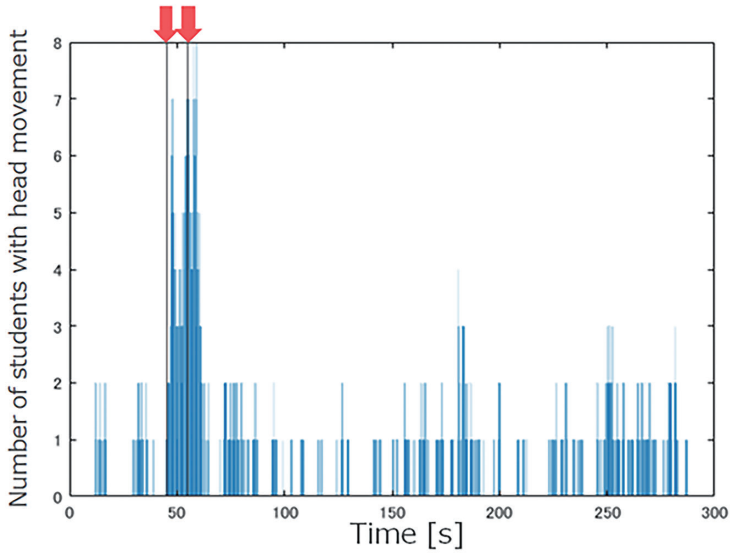

Figure 3. Total number of students who moved their heads each second in a group based on the camera-sensor data. Arrows indicate a large number of students moving their heads at the same time. [The figure was created by the authors based on Haruta et al.'s study's study (2020).]

The observation revealed the extent to which students were looking at the guide or the object, and how focused they were on the guide's talk. Figure 4 shows an example of four students with their scores. The mean of the seven scores (with a total of 210 seconds observation scored every 30 seconds) on their eye movements was calculated as their overall observation score (Table 1). Numbers such as D-2 in Figure 3 represent the IDs of the students and correspond to the students with the same ID numbers as Table 1.

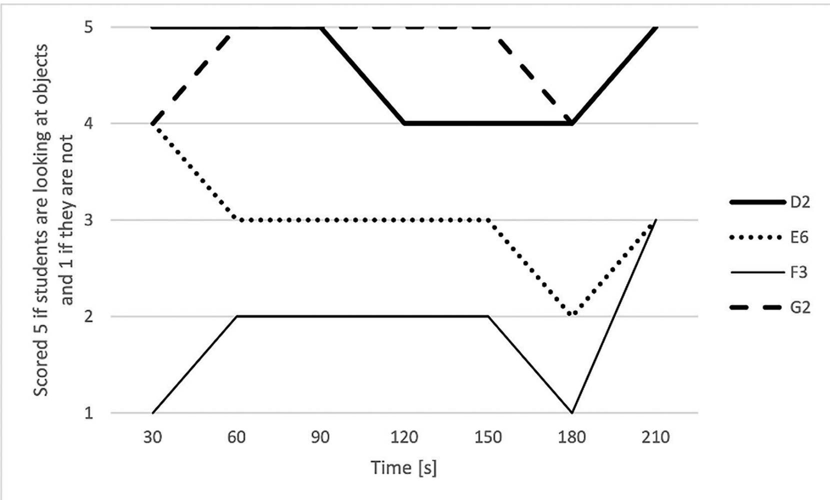

Figure 4. The degree to which students looked at the guide or the object, based on observation. D2, E6, F3, and G2 represent the IDs of the students. (Created by the authors)

Table 1. Comparison of scores obtained from cameras, observation, interviews, and self-reported surveys. (Created by the authors)

Student ID	Camera/ sensor score	Observation score	If student remembered guide's talk	Self-reported score of focus [Interview]	Self-reported score of enjoyment [Interview]	Self-reported score of focus [Survey]	Self-reported score of interest [Survey]
A4	-27.7	3.21	1	90	100	4.5	3.5
B2	229	4.79	0	70	100	5.0	5.0
C5	56.1	3.79	1	38	100	3.5	4.0
D2	-181	4.57	1	85	100	4.5	5.0
E6	-58	3.36	1	90	100	5.0	4.5
F3	-551	2.00	1	50	90	3.0	3.5
G2	-42.5	4.79	0	60	80	5.0	4.5

The results of the interviews with the seven students are shown in Table 1. If students remembered the guide's talk well, their score was 1 (e.g., "The guide was talking about invasive species and how these species enter Shiga Kogen by attaching to people's clothes and shoes"), and if not (e.g., "I don't remember"), their score was 0. In addition, students' self-reported scores of their own performance are shown in Table 1.

The results of the surveys conducted after the tour are shown as "self-reported score of focus" which are mean scores of students on statements "I listened carefully to the guide's talk from beginning to end" and "I looked at and listened to what the guide explained and pointed at" in Table 1. Mean scores of statements "I listened to the guide's talk with interest" and "It was fun to listen to the guide's talk" are shown as "self-reported scores of interest."

(2) Comparison of Data Obtained through Different Methods of Evaluation

The student with the highest score for the camera/sensor (B2 = 229) reacted fastest to the guide's talk and secured the highest score for observation (4.79), while the student with the lowest score for the camera/sensor (F3 = -551) also had the lowest score for observation (2.00). We divided the students into two groups: those who had negative scores for camera/sensor and those who had positive scores for camera/sensor. We also divided the students into two categories: those having low scores for observation (<3.50) and those who had high scores for observation ($\geqq 3.50$). The scores of five out of seven students on evaluation through camera/sensor and observation (e.g., negative camera scores and low observation scores), were similar. The students D2 and G2 displayed contradictory results (i.e., both had high scores on observation and low scores on camera/sensor). Therefore, from these results, camera/sensor monitoring appears to generate similar results as human observation for most of the samples. There was a medium correlation between camera and observation scores although it was not statistically significant (Spearman correlation = 0.487; $p = 0.268$) (Table 2).

Table 2. Correlation matrix based on Spearman's rank correlation coefficient. (Created by the authors)

	Camera/ sensor score	Observation score	If student remembered guide's talk	Focus [Interview]	Enjoyment [Interview]	Focus [Survey]	Interest [Survey]
Camera/ sensor score	1.000	0.487	-0.474	-0.072	0.401	0.356	0.184
Observation score	0.847	1.000	-0.798*	-0.109	-0.045	0.661	0.815*
If student remembered guide's talk	-0.474	-0.798*	1.000	0.160	0.394	-0.663	-0.488
Focus [Interview]	-0.072	-0.109	0.160	1.000	0.449	0.500	0.185
Enjoyment [Interview]	0.401	-0.045	0.394	0.449	1.000	0.047	0.252
Focus [Survey]	0.356	0.661	-0.663	0.500	0.047	1.000	0.635
Interest [Survey]	0.184	0.815*	-0.488	0.185	0.252	0.635	1.000

(*: p,.05)

The student who secured the highest score for both camera/sensor and observation scored highly (score = 5.0) on the self-reported survey measuring their focus, while the student who secured the lowest score on camera/sensor and observation scored low on the self-reported survey too (score = 3.0). This shows that results of camera/sensor, observation and that of the survey were similar. However, the scores of other students (A4 and E6) are contradictory as those students who scored low for camera/sensor and observation scored high for their survey (4.5 and 5.0, respectively). Contradictory results were also obtained for how much they remembered about the guide's talk as well as for their self-reported scores. The student with the highest score on camera/sensor and observation (B2) did not remember the contents of the guide's talk while the student who had the lowest score (F3) was able to articulate the contents of the guide's talk. While the student with the lowest score on camera/sensor and observation also scored low for focus on the guide's talk in the interview (score = 50), the student with the highest score (B2) as well as the student who had a fairly high score (C5) for camera/sensor and observation also obtained a low score on the interview (i.e., 70 and 38, respectively).

Correlation analysis among the seven factors [(1) camera/sensor score, (2) observation score, (3) knowledge score, (4) self-reported score of focus by interview, (5) self-reported score of enjoyment by interview, (6) self-reported score of focus by survey, and (7) self-reported score of interest by survey] (Table 1) revealed that there were two statistically significant correlations. Observation scores correlated with students' knowledge (if the student remembered the guide's talk) (Spearman correlation = -0.798; p = 0.032) as well as with self-reported interest scores (Spearman correlation = 0.815; p = 0.026) as per the survey answers. This not only shows the potential significance but also the limitation of researchers' observation. Those students who had high observation scores also scored high in self-reported interest, however, they had lower knowledge scores.

Our research also suffered from the limitation of self-reported surveys and interviews. In our study, student (C-5) scored well at paying attention and was able to explain the contents of the guide's talk better compared to the other students, while she scored herself low on attention in the survey. Differences between scores measured in self-reported surveys and what others (such as observers) pointed out were also seen in previous studies (e.g., Lee et al. 2018). A limitation of self-reported studies is that participants may score themselves low if they are humble. Conversely, opposite results can be obtained if participants are too confident and score themselves higher than their actual performance. Meanwhile, the fact that observation scores statistically corresponded with students' self-reported scores of interest in the survey implies that the degree to which the students were looking at the objects (shown from observation) was reflected in their level of interest in the guide's talk and the tour.

Although the correlation between scores of camera and observation was moderate and not significant, we might see stronger association by increasing the sample size or by excluding the outliers (data point that differs significantly from other observation). For example, assuming that the two students D2 and G2 (who scored low in camera data and high in observation) were outliers, correlation between camera and observation scores became high and significant (Spearman correlation = 0.900; $p = 0.037$). Therefore, camera and sensor devices could potentially be substituted for human observation in the future, if we are able to articulate and exclude outliers from the samples. Meanwhile, even after deleting two samples (D2 and G2), camera scores did not show any significant correlation with the students' scores on the interview and the survey. This implies that camera and sensor technology has the potential to substitute researchers' observation but not the participants' perceptions. In other words, the objective evaluation that we applied was not effective in terms of depicting participants' cognitive elements, and posed concerns that objective evaluation, or at least the approach we developed and utilized in this research, was not necessarily "correct evaluation." Further research is necessary to understand whether camera and observation can only be used for recording participants' reactions, or whether such technology and observation are more effective in procuring data than self-reported measures for determining cognitive outcomes.

(3) Advantage of Evaluation Using Camera/Sensor over Conventional Social Science Methods

One advantage of a camera/sensor is that it can record and measure head movement at the same time and provide objective outcomes as quantitative data. Subjective evaluation methods that involve people (e.g., participants, observers) are usually expensive and time-consuming (Camargo and Shavelson 2009; Pourmand 2012). The reason why only one student from each group was monitored for observation was that we had only one researcher available to observe students. Hiring enough researchers or assistants to observe all participants is challenging both financially and logistically (Camargo and Shavelson 2009). Creating machines with the help of engineers could provide additional information for conventional survey, observation or interview research and could overcome the issues of subjectivity as well as diminish the necessity of hiring many assistants. In addition, as the quality of interviews and observations could depend on the skills of interviewers and observers (e.g., level of training), interviewers could generate different outcomes (Brinkmann and Kvale 2015, p. 71)]. The results of objective evaluation (in this study, use of camera and censors) are consistent and repeatable, and data obtained from different studies and locations could be directly compared (Pourmand 2012).

Camera/sensor data provide detailed and objective information related to participants' physical reactions during the program (e.g., the moment participants started to lose attention) to the environmental educators as well as tour guides (interpreters), and by utilizing such data, practitioners could design and conduct more effective environmental education and communication programs. In that sense, obtaining data by machines could be useful for evaluating not only environmental education programs but also the guides (environmental educators) and how well they were able to attract respondents' attention during the guided tour. This is especially important since many environment educators tend to assume that participants have the same motivation toward the programs as they do and conducting an evaluation is the only way to reveal such gaps (Heimlich 2010).

5. Conclusion

The field of EE evaluation has been developed mainly by social scientists who use social science approaches, such as surveys and interviews, and social psychological theories (Camargo and Shavelson 2009; Zint 2013). Indeed, these disciplines have made important contributions to the development of the EE field. Our idea of creating machines to objectively and systematically evaluate the impact of EE programs augments traditional methods and may provide a new paradigm for the field of EE research. Based on our results, new academic and interdisciplinary theories could be developed on the relationship between subjective and objective evaluations. In other words, the implications of our study results are that collaboration with engineers could potentially challenge the existing paradigm, move the field forward to the next stage of EE evaluation, and provide an alternative research model for overcoming the limitations of conventional evaluation methods. However, a limitation of this study was that, firstly, it was a pilot study with limited sample size. A large scale study that yields more extensive data is necessary to confirm the credibility of our approach. Secondly, experimental research should be conducted for a more rigorous assessment of the evaluation tools. Wearing attachable cameras could have affected students' reaction during the tour, so future research can be conducted by dividing participants into treatment and control groups to investigate the effects of such experiments. Thirdly, assuming that head movement itself cannot identify the full mechanism of students' learning process, we need to keep exploring and considering the meaning of the data obtained from such wearable cameras. The objective of this research; to identify whether camera/sensor data could substitute conventional observation, was partly verified by our data. However, as the next step, we should go beyond just testing whether objective evaluation could substitute conventional observation, and explore what aspects of the learning process could be evaluated through such camera/sensor data.

Finally, since the ultimate goal of environmental education is to increase citizen's awareness toward sustainability as well as their pro-environmental behavior, future research should consider how such subjective and objective evaluation could potentially identify programs' impacts. As for the next step, creating and utilizing rubrics that identify the objective criteria of the program's impacts (e.g., American Association of Colleges and Universities 2022) as well as conducting peer evaluation and third-party evaluation could be important to bridge the gap between our findings (e.g., students' reactions) and the program's impacts (e.g., increase in pro-environmental behavior).

References

American Association of Colleges and Universities. 2022. VALUE Rubrics: Civic Engagement. <https://www.aacu.org/initiatives/value-initiative/value-rubrics/value-rubrics-civic-engagement> (accessed December 23, 2022)

Asch, J., and B. M. Shore. 1975. Conservation Behavior as the Outcome of Environmental Education. *The Journal of Environmental Education*, 6(4), 25–33.

Ali, S. B., and S. Patnaik. 2018. Thermal Comfort in Urban Open Spaces: Objective Assessment and Subjective Perception Study in Tropical City of Bhopal, India. *Urban Climate*, 24, 954–967.

Brinkmann, S., and S. Kvale. 2015. *Interviews: Learning the Craft of Qualitative Research Interviewing*. 3rd ed. California: Sage.

Camargo, C., and R. Shavelson. 2009. Direct Measures in Environmental Education Evaluation: Behavioral Intentions Versus Observable Actions. *Applied Environmental Education and Communication*, 8(3-4), 165173.

Duvall, J., and M. Zint. 2007. A Review of Research on the Effectiveness of Environmental Education in Promoting Intergenerational Learning. *The Journal of Environmental Education*, 38(4), 14–24.

Ernst, J. A., M. Monroe, and B. Simmons. 2009. *Evaluating Your Environmental Education Programs: A Workbook for Practitioners*. Washington, D. C.: North American Association for Environmental Education (NAAEE).

Haruta, K., K. Hashimoto, R. Sakurai, K. W. Nakamura, and T. Nakata. 2020. Examination of Group Head Angle Acceleration Analysis Method for Learning Evaluation in Outdoor Education. In *International Workshop on Advanced Image Technology (IWAIT 2020)*, (Volume 11515, p. 1151518) International Society for Optics and Photonics.

Heimlich, J. E. 2010. Environmental Education Evaluation: Reinterpreting Education as a Strategy for Meeting Mission. *Evaluation and Program Planning*, 33(2), 180–185.

Jacobson, S. K., M. D. McDuff, and M. C. Monroe. 2007. *Conservation Education and Outreach Techniques*. Oxford: Oxford University Press.

Kasahara, S., S. Nagai, and J. Rekimoto. 2017. JackIn Head: Immersive Visual Telepresence System with Omnidirectional Wearable Camera. *IEEE Transactions on Visualization and Computer Graphics*, 23(3), 1222–1234.

Kollmuss, A., and J. Agyeman. 2002. Mind the Gap: Why do People Act Environmentally and What are the Barriers to Pro-Environmental Behavior? *Environmental Education Research*, 8(3), 239–260.

Kormos, C., and R. Gifford. 2014. The Validity of Self-Report Measures of Proenvironmental Behavior: A Meta-Analytic Review. Journal of Environmental Psychology, 40, 359–371.

Lee, K., C. Yu, P. Hsieh, C. Li, and Y. C. Chao. 2018. Situated Teaching Improves Empathy Learning of the Students in a BSN Program: A Quasi-Experimental Study. *Nurse Education Today*, 64, 138–143.

Lusk, J. L., and F. B. Norwood. 2009. An Inferred Valuation Method. *Land Economics*, 85(3), 500–514.

Nishimura, H. 2016. Interpretation Katudo no Atarashii Doko (The New Trends of the Interpretive Activities). *Doshisha University policy & management review*, 20th Year Anniversary Special Issue, 91–98. (In Japanese).

Onishi, A., K. Saito, and T. Terada. 2016. Sochaku-gata Sensa wo Mochiita Yomikikase-ji no Kyomi Suitei Shuho no Hyoka (Recognizing Behavior and Predicting Interests During Reading Time Using Wearable Sensor). *Proceedings of Multimedia, Distributed, Cooperative, and Mobile Symposium*. 1175–1182. (In Japanese).

Pourmand, N. 2012. O*bjective and Subjective Evaluation of Wideband Speech Quality*. [Doctoral dissertation, The University of Western Ontario]. The University of Western Ontario Theses and Dissertations Archive.

Rossi, P. H., M. W. Lipsey, and H. E. Freeman. 2004. *Evaluation: a Systematic Approach*. Seventh Edition. California: Sage Publications, Inc.

Rothkopf, C. A., and J. B. Pelz. 2004. Head Movement Estimation for Wearable Eye Tracker. *Proceedings of the 2004 Symposium on Eye Tracking Research & Applications*, 123–130.

Vaske, J. J. 2008. *Survey Research and Analysis: Applications in Parks, Recreation, and Human Dimensions*. Pennsylvania: Venture Publishing Inc.

Velarde, M. D., G. Fry, and M. Tveit. 2007. Health Effects of Viewing Landscapes – Landscape Types in Environmental Psychology. *Urban Forestry & Urban Greening*, 6(4), 199–212.

Zint, M. 2013. Chapter 30. Advancing Environmental Education Program Evaluation: Insights from a Review of Behavioral Outcome Evaluations. In R. B. Stevenson, M. Brody, J. Dillon, and A. E. J. Wals (eds). *International Handbook of Research on Environmental Education*, New York: Routledge, 298–309.

Appendix 1. Equation used to Evaluate Head Direction Movement.

Based on the following equation, we evaluated if the speed of head direction movement surpassed the threshold.

$$if(|v_t| > \alpha\delta)$$

We counted the number of students who moved their heads: $if(s_t > 3/N)$,

and calculated student's individual scores: $E{+}{=}\ |v_t|$, $elseif(s_t <= 3/N)$, $E{-}{=}\ |v_t|$

t refers to time, v_t refers to head direction speed, s_t refers to the number of students who moved their heads, *N* refers to the total number of people in the group, and *E* refers to the individual scores of the students.